

## Article

# Novel Structural Variation and Evolutionary Characteristics of Chloroplast tRNA in *Gossypium* Plants

Ting-Ting Zhang<sup>1</sup>, Yang Yang<sup>1</sup>, Xiao-Yu Song<sup>1</sup>, Xin-Yu Gao<sup>1</sup>, Xian-Liang Zhang<sup>2</sup>, Jun-Jie Zhao<sup>2</sup>, Ke-Hai Zhou<sup>2</sup>, Chang-Bao Zhao<sup>2</sup>, Wei Li<sup>2</sup>, Dai-Gang Yang<sup>2</sup>, Xiong-Feng Ma<sup>2,3,\*</sup> and Zhong-Hu Li<sup>1,\*</sup>

- <sup>1</sup> Key Laboratory of Resource Biology and Biotechnology in Western China, Ministry of Education, College of Life Sciences, Northwest University, Xi'an 710069, China; zhangtt@stumail.nwu.edu.cn (T.-T.Z.); 202021088@stumail.nwu.edu.cn (Y.Y.); songxiaoyu@stumail.nwu.edu.cn (X.-Y.S.); 202032597@stumail.nwu.edu.cn (X.-Y.G.)
- <sup>2</sup> State Key Laboratory of Cotton Biology, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang 455000, China; zhangxianliang@caas.cn (X.-L.Z.); zhaojunjie@caas.cn (J.-J.Z.); zhoukehaid@caas.cn (K.-H.Z.); zhaochagnbao@caas.cn (C.-B.Z.); liwei@caas.cn (W.L.); yangdaigang@caas.cn (D.-G.Y.)
- <sup>3</sup> School of Agricultural Sciences, Zhengzhou University, Zhengzhou 450001, China
- \* Correspondence: maxiongfeng@caas.cn (X.-F.M.); lizhonghu@nwu.edu.cn (Z.-H.L.); Tel./Fax: +86-29-88302411 (Z.-H.L.)



**Citation:** Zhang, T.-T.; Yang, Y.; Song, X.-Y.; Gao, X.-Y.; Zhang, X.-L.; Zhao, J.-J.; Zhou, K.-H.; Zhao, C.-B.; Li, W.; Yang, D.-G.; et al. Novel Structural Variation and Evolutionary Characteristics of Chloroplast tRNA in *Gossypium* Plants. *Genes* **2021**, *12*, 822. <https://doi.org/10.3390/genes12060822>

Academic Editor:  
Oren Ostersehtzer-Biran

Received: 29 March 2021  
Accepted: 24 May 2021  
Published: 27 May 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Cotton is one of the most important fiber and oil crops in the world. Chloroplast genomes harbor their own genetic materials and are considered to be highly conserved. Transfer RNAs (tRNAs) act as “bridges” in protein synthesis by carrying amino acids. Currently, the variation and evolutionary characteristics of tRNAs in the cotton chloroplast genome are poorly understood. Here, we analyzed the structural variation and evolution of chloroplast tRNA (cp tRNA) based on eight diploid and two allotetraploid cotton species. We also investigated the nucleotide evolution of chloroplast genomes in cotton species. We found that cp tRNAs in cotton encoded 36 or 37 tRNAs, and 28 or 29 anti-codon types with lengths ranging from 60 to 93 nucleotides. Cotton chloroplast tRNA sequences possessed specific conservation and, in particular, the Ψ-loop contained the conserved U-U-C-X3-U. The cp tRNAs of *Gossypium* L. contained introns, and cp tRNA<sup>Ile</sup> contained the anti-codon (C-A-U), which was generally the anti-codon of tRNA<sup>Met</sup>. The transition and transversion analyses showed that cp tRNAs in cotton species were iso-acceptor specific and had undergone unequal rates of evolution. The intergenic region was more variable than coding regions, and non-synonymous mutations have been fixed in cotton cp genomes. On the other hand, phylogeny analyses indicated that cp tRNAs of cotton were derived from several inferred ancestors with greater gene duplications. This study provides new insights into the structural variation and evolution of chloroplast tRNAs in cotton plants. Our findings could contribute to understanding the detailed characteristics and evolutionary variation of the tRNA family.

**Keywords:** cotton; chloroplast tRNA; evolution; phylogenetic relationship; structural variation

## 1. Introduction

Chloroplasts (cp) are unique oblate organelles equipped with important photosynthesis functions [1]. Besides the complete photosynthetic components, they have also been reported as active organelles participating in other diverse biomolecule processes including biosynthesis of starch, fatty acids, pigments, amino acids, etc. [2–4]. Some previous studies found that chloroplasts originated from cyanobacterial ancestors [5,6]. For most plants, the cp genome is a double-stranded circular unit involving four main regions: two inverted repeats (IRs) separated by a large single copy (LSC) region and a small single copy (SSC) region [7,8]. In addition, the features of non-recombination and maternal inheritance (in most angiosperms) of the cp genome provide the opportunity to employ it as material for

evolutionary and genomics research [9,10]. There are many genes, such as protein-coding genes and multiple tRNA genes, present in the cp genome [11]. Relevant studies have shown that most of these genes play a role in the organization of photosynthesis and biochemical reactions [12–14]. However, it is worth noting that the genetic and mutational characteristics of cp genes, especially cp tRNA genes in angiosperms, still need research.

Generally, single nucleotide polymorphisms (SNPs) and insertions/deletions are the basis of the differences between most alleles. With the characteristics of high abundance, a fairly low mutation rate, and the adaptability to automatic genotyping, SNPs are employed more frequently than other genetic markers, such as microsatellites [15]. Insertions and deletions (indels) are essential sources of polymorphic markers for high-resolution genetic mapping of traits and association studies based on candidate genes or possibly the whole genome [16]. Furthermore, SNPs and indels are also important in the nucleotide sequence evolution analysis of genomes. The calculation of single nucleotide variation may help estimate and comprehend genetic variations of different genome regions.

According to previous studies, the residue polynucleotide sequences of tRNAs fold back, forming clover leaf-like structures with hydrogen bonds, and then turn into L-shaped tertiary structures [17]. The secondary structure of tRNA contains an acceptor arm, dihydrouridine arm (D-arm), dihydrouridine loop (D-loop), anti-codon arm, anti-codon loop, variable loop, pseudouridine arm ( $\Psi$ -arm), and pseudouridine loop ( $\Psi$ -loop) [18,19]. With this specific structure, tRNA plays an important role in protein synthesis by carrying amino acids to the ribosome [20,21]. Recently, researchers conducted a series of analyses related to tRNAs on the genomic level. Asymmetric combinations and divided segments in tRNA genes would help to understand the diversity of tRNA molecules [22]. Additionally, studies showed that the heterogeneous tRNA fragments play multiple roles in terms of size, nucleotide composition, biogenesis, and even biological disease [23]. Additionally, wobble modifications were frequently found in tRNAs in diverse species after the discovery of tRNA molecules [24–26]. In recent years, studies on the genomic structure and evolution of tRNAs in plants has been attracting researchers' attention. For example, analyses of detailed molecular aspects in cyanobacterial tRNAs, complete genomic features of tRNAs in *Oryza sativa* L., and the evolutionary perspective of chloroplast tRNAs (cp tRNAs) in some economic monocots were previously conducted [27–29].

*Gossypium* L., or the cotton plant, is an important economic and oil crop that has been cultivated worldwide. *Gossypium* is a large genus belonging to the angiosperm family Malvaceae. Fryxell divided the cotton genus into four subgenera, with a total of 51 species, including 46 diploid ( $2n = 2 \times = 26$ ) and five tetraploid ( $2n = 4 \times = 52$ ) species [30]. The chromosome composition of tetraploid cotton is heterogeneous A and D: *Gossypium hirsutum* L. (AD1), *G. barbadense* L. (AD2), *G. tomentosum* Nuttalex Seemann (AD3), *G. mustelinum* Miersex Watt (AD4), and *G. darwinii* Watt (AD5). A total of eight chromosomes of diploid cotton belong to groups A–G and K. Among them, *G. hirsutum*, *G. barbadense*, *G. arboreum* L., and *G. herbaceum* L. are cultivars, while the others are wild species. Nowadays, the most widely cultivated cotton species are tetraploid *G. hirsutum* and *G. barbadense* [31]. At present, with the rapid development of genome sequencing technology, whole-genome sequences of cotton have been released, which provide the foundation for further analysis of cotton genomes [32,33]. The genome-wide landscape of genomic variation of cotton was constructed through SNP distribution density detection, and functional genes that encode proteins involved in regulation of tissue growth, stress responses, and disease resistance were reported [34–39]. Moreover, a previous study on variations of repeat sequences and cp evolutionary relationships detected divergence hotspots in plastid genomes and helped to understand phylogenetic relationships among major *Gossypium* lineages [40]. However, the evolutionary patterns of *Gossypium* chloroplast tRNAs are still unclear. The study of the genomic and evolutionary characteristics of cotton cp tRNAs seems to be significant.

In this study, we investigated ten globally representative cotton cp genomes, including eight diploid and two allotetraploid species. The aims of our study were: (1) identify the

genomic characteristics and diversification of cp tRNAs in cotton; (2) analyze the evolutionary relationship of introns in cp tRNA genes; (3) estimate the evolutionary characteristics of SNPs and the indel mutation rate of the cotton chloroplast genome; and (4) investigate the evolutionary pattern of cp tRNAs in cotton species.

## 2. Materials and Methods

### 2.1. Identification of tRNAs

We downloaded the 10 cotton cp genomes (8 diploid and 2 allotetraploid species: *Gossypium arboreum* L., *G. anomalum* Wawra and Peyritsch, *G. robinsonii* (F. Muell.) J. H. Willis, *G. klotzschianum* Andersson, *G. somalense* (Gurke) J. B. Hutch., *G. longicalyx* Hutchinson and Lee, *G. bickii* Prokhanov, *G. populifolium* (Benth.) F. Muell., *G. hirsutum* L., and *G. barbadense* L.) from the National Center of Biotechnology Information (NCBI) (Table 1). These cotton species are widely distributed near the equator and *G. hirsutum*, *G. barbadense*, and *G. klotzschianum* are mainly distributed in the Americas. Subsequently, tRNA gene sequences were identified and extracted from cp genomes without intergenic regions by the GENEIOUS 8.0.2. program [41].

**Table 1.** *Gossypium* species information in the study.

Karyotype	Species	Accession Number	Wild/Cultivars
A2	<i>Gossypium arboreum</i>	NC_016712	cultivars
B1	<i>G. anomalum</i>	NC_023213	wild
C2	<i>G. robinsonii</i>	NC_018113	wild
D3-k	<i>G. klotzschianum</i>	NC_033394	wild
E2	<i>G. somalense</i>	NC_018110	wild
F1	<i>G. longicalyx</i>	NC_023216	wild
G1	<i>G. bickii</i>	NC_023214	wild
K2	<i>G. populifolium</i>	NC_033398	wild
AD1	<i>G. hirsutum</i>	HQ901196	cultivars
AD2	<i>G. barbadense</i>	HQ901199	cultivars

### 2.2. Structural Analysis of tRNAs

ARAGORN and tRNAScan-SE software [42,43] were employed to investigate the secondary structure of tRNA sequences of cp genomes. The default parameters of ARAGORN software were set to investigate tRNAs. The parameters of tRNAScan-SE were set as: bacterial for sequence source, default for search mode, formatted (FASTA) for query sequences, and universal for genetic code for tRNA isotype prediction.

### 2.3. Sequence Alignment

To identify the presence of consensus sequences, sequence alignments were carried out for the intron sequences of the cotton cp tRNAs, *Pinus armandii* Franch., *Marchantia polymorpha* L., *Raphanus sativus* L., *Spirogyra maxima* (Hassall) Wittrock, *Alsophila spinulosa* (Wallich ex Hooker) R. M. Tryon Contr. Gray Herb., *Zea mays* L., *Gleocapsa* sp. PCC 73106, *Nostoc* sp. PCC 7107, and *Nostoc* sp. PCC 7524 using Multalin software, in which default parameters were set [44–46].

### 2.4. Phylogenetic Tree Construction

The phylogenetic tree was constructed using MEGA7.0 software [47,48]. To investigate the evolution of chloroplast tRNAs, a matrix of whole tRNA sequences was created by Clustal Omega software before the phylogenetic tree was constructed. MEGA7 software was employed to turn the matrix file into MEGA file format for tree construction, in which the lowest Bayesian information criterion (BIC) was selected for the model. As a result, the Kimura2 + G + I model was found to have the lowest BIC score, 8980.50. Thus, this model was adopted for the phylogenetic tree construction. The other related parameters were as follows: phylogeny reconstruction for analysis, maximum likelihood model, bootstrap method in phylogeny test, 1000 bootstrap replicates, nucleotide type,  $\gamma$  distributed with

invariant sites (G + I) model, 5 discrete  $\gamma$  categories, partial deletion for gaps/missing data treatment, 95% site coverage cutoff, and very strong for branch swap filter.

### 2.5. Analysis of Disparity Index

A disparity index test of pattern heterogeneity was conducted to check the homogeneity of nucleotide substitutions and find whether all substitutions in nucleotides happened at equal rates, i.e., homogeneity in the process of evolution. The statistical parameters were as follows based on a previous study [29]: disparity index test for substitution pattern homogeneity, in sequence pairs, 10,001 Monte Carlo replications, nucleotide for substitution type, partial deletion of gaps/missing data treatment, and site coverage cutoff 95%.

### 2.6. Transition/Transversion Analysis

The transition and transversion rates of the tRNAs genes were analyzed according to their isotypes by MEGA 7 software [49]. The parameters were set as follows: substitution pattern estimation (ML) for analysis, automatic (neighbor-joining tree), maximum likelihood statistical method, nucleotide for substitution type, Kimura 2-parameter model,  $\gamma$  distributed (G) site rates, 5 discrete  $\gamma$  categories, partial deletion of gaps/missing data treatment, 95% of site coverage cutoff, and very strong branch swap filter.

### 2.7. Evolutionary Analysis of Single Nucleotide Polymorphisms

The diversity of single nucleotide polymorphisms of 8 diploid *Gossypium* cp genomes was calculated based on 3 regions: coding regions, introns, and intergenic spacers. The synonymous (dS) and non-synonymous (dN) substitutions of coding genes were also calculated by DnaSP v5.10 software [50]. The coding regions, introns, and intergenic regions of the studied *Gossypium* cp genome were extracted through Geneious 8.0.2 software and aligned manually [41].

### 2.8. Calculation of Mutation Rate

The rate of indel mutation ( $\mu$ , per site per year) was calculated with the formula:

$$\mu = m/(nT), \quad (1)$$

where  $m$  is the number of sites of observed mutation,  $n$  is the total number of sites, and  $T$  is the divergence time of *Gossypium*. The  $\mu$  value of structural mutations was calculated according to the method of Saitou and Ueda [51]. The  $T$  value was obtained through relevant published literature searches in the Fossil works database and the Cenozoic Angiosperm Database [52,53]. Additionally, the mutation rates of protein-coding genes and tRNA genes were calculated.

### 2.9. Duplication/Loss Analysis of tRNA Genes

To investigate duplication or loss events of the tRNAs, we used the NCBI taxonomy browser to construct the species tree of *G. arboreum*, *G. anomalum*, *G. robinsonii*, *G. klotzschianum*, *G. somalense*, *G. longicalyx*, *G. bickii*, *G. hirsutum*, *G. barbadense*, and *G. populifolium*. Additionally, the previously constructed phylogenetic tree of the tRNAs was employed as the gene tree. Subsequently, Notung 2.9 software [54,55] was used to reconcile the gene tree and the species tree and obtain the gene duplication and loss nodes.

## 3. Results

### 3.1. Basic Characteristics of Cotton Chloroplast tRNAs

The results of detailed genomic analysis showed that *G. arboreum*, *G. anomalum*, *G. robinsonii*, *G. klotzschianum*, *G. somalense*, *G. hirsutum*, *G. barbadense*, *G. longicalyx*, and *G. populifolium* coded 37 tRNAs, respectively, while only *G. bickii* coded 36 tRNAs (Table 2). The length of the chloroplast tRNAs ranged from 60 nt (tRNA<sup>Gly</sup> in *G. arboreum*, GCC) to 93 nt (tRNA<sup>Ser</sup>, UGA), with an average length of 76 nt (Table S1).

**Table 2.** Distribution of tRNA isotypes in cotton chloroplast genome.

tRNA Isotype	Number of tRNAs									
	A2 <sup>1</sup>	B1 <sup>2</sup>	C2 <sup>3</sup>	D3-K <sup>4</sup>	E2 <sup>5</sup>	F1 <sup>6</sup>	AD1 <sup>7</sup>	AD2 <sup>8</sup>	G1 <sup>9</sup>	K2 <sup>10</sup>
Alanine	2	2	2	2	2	2	2	2	2	2
Glycine	2	2	2	2	2	2	2	2	2	2
Proline	1	1	1	1	1	1	1	1	1	1
Threonine	2	2	2	2	2	2	2	2	2	2
Valine	3	3	3	3	3	3	3	3	3	3
Serine	3	3	3	3	3	3	3	3	3	3
Arginine	3	3	3	3	3	3	3	3	3	3
Leucine	4	4	4	4	4	4	4	4	4	4
Phenylalanine	1	1	1	1	1	1	1	1	1	1
Asparagine	2	2	2	2	2	2	2	2	2	2
Lysine	1	1	1	1	1	1	1	1	1	1
Aspartate	1	1	1	1	1	1	1	1	1	1
Glutamate	1	1	1	1	1	1	1	1	1	1
Histidine	1	1	1	1	1	1	1	1	1	1
Glutamine	1	1	1	1	1	1	1	1	1	1
Isoleucine	4	4	4	4	4	4	4	4	4	4
Methionine	2	2	2	2	2	2	2	2	1	2
Tyrosine	1	1	1	1	1	1	1	1	1	1
Cysteine	1	1	1	1	1	1	1	1	1	1
Tryptophan	1	1	1	1	1	1	1	1	1	1
Selenocysteine	0	0	0	0	0	0	0	0	0	0
Suppressor	0	0	0	0	0	0	0	0	0	0
Total	37	37	37	37	37	37	37	37	36	37

<sup>1</sup> *Gossypium arboreum*; <sup>2</sup> *G. anomalum*; <sup>3</sup> *G. robinsonii*; <sup>4</sup> *G. klotzschianum*; <sup>5</sup> *G. somalense*; <sup>6</sup> *G. longicalyx*; <sup>7</sup> *G. hirsutum*; <sup>8</sup> *G. barbadense*; <sup>9</sup> *G. bickii*;

<sup>10</sup> *G. populifolium*.

The genomic analysis results showed that chloroplast tRNA genes of the investigated cotton plants coded 28 or 29 anti-codon types, of which only *G. anomalum*, *G. longicalyx*, and *G. bickii* coded 29 anti-codons. The most common anti-codons observed in cp tRNAs were UGC-tRNA<sup>Ala</sup>, GCC-tRNA<sup>Gly</sup>, GAC-tRNA<sup>Val</sup>, ACG-tRNA<sup>Arg</sup>, CAA-tRNA<sup>Leu</sup>, GUU-tRNA<sup>Asn</sup>, CAU-tRNA<sup>Ile</sup>, GAU-tRNA<sup>Ile</sup>, and CAU-tRNA<sup>Met</sup>. Additionally, each of these genes, except GCC-tRNA<sup>Gly</sup> and CAU-tRNA<sup>Met</sup>, had two copies (Table S2). GCC (tRNA<sup>Gly</sup>) was found with one copy in *G. anomalum*, *G. longicalyx*, and *G. bickii* but two copies in the other cotton species. Similarly, CAU (tRNA<sup>Met</sup>) was found with one copy in *G. bickii* but two copies in the other cotton species. There were 33 tRNAs with various iso-acceptors missing in the cp genome of *Gossypium* (Table S2). UCC-tRNA<sup>Gly</sup> was observed in *G. anomalum*, *G. longicalyx*, and *G. bickii*. All investigated cotton species were shown to possess at least one anti-codon type for each kind of tRNA. Furthermore, the anti-codon CAU was a typical characteristic of tRNA<sup>Met</sup>, which harbored only one type of iso-acceptor. Apart from the existence of the anti-codon CAU in tRNA<sup>Met</sup>, tRNA<sup>Ile</sup> was also observed to encode the anti-codon CAU in the cotton cp genome (Table S2).

All the tRNA gene families were analyzed by multiple sequence alignment, from which the limited conservation and consensus sequences were found in the Ψ-arm and Ψ-loop. The Ψ-arm of tRNAs was observed to contain the G-G consensus sequence, and the Ψ-loop was observed to contain the conserved sequence U-U-C-X3-U (Table 3).

Most of the tRNAs possessed a G nucleotide at the first position of the acceptor arm, while tRNA<sup>Gln</sup>, tRNA<sup>Pro</sup>, and tRNA<sup>Val</sup> were observed to contain a U and an A (Table 3). Except for tRNA<sup>Lys</sup>, tRNA<sup>Met</sup>, tRNA<sup>Thr</sup>, tRNA<sup>Tyr</sup>, and tRNA<sup>Val</sup>, other tRNAs contained a G in the first nucleotide of the D-arm. Additionally, there was usually an A at the last position of the D-loop and ANC-loop. At the final site of the D-arm, except for tRNA<sup>Arg</sup>, tRNA<sup>Glu</sup>, tRNA<sup>Ile</sup>, tRNA<sup>Leu</sup>, tRNA<sup>Met</sup>, tRNA<sup>Ser</sup>, and tRNA<sup>Tyr</sup>, the other investigated tRNAs were found to have a C (Table 3). Additionally, we also observed that all tRNA<sup>Lys</sup>, tRNA<sup>Ala</sup>, tRNA<sup>Glu</sup>, tRNA<sup>Arg</sup>, tRNA<sup>Tyr</sup>, and a few tRNA<sup>Leu</sup> had a C-C-A tail in the 3' end.



**Table 3.** Sequence alignment and the presence of isotype-specific conserved nucleotide consensus sequences in the cotton chloroplast tRNAs.

tRNA Isotype	AC-Arm	D-Arm	D-Loop	ANC-Arm	ANC-Loop	Variable Region	Ψ-Arm	Ψ-Loop
Alanine	GGGGAUA	GCUC	AGUUGGUA	CCGCU	CUUGCAU	AUGUC	AGCGG	UUCGAGU
Arginine	GXGXC <sub>2</sub>	Gx <sub>3</sub>	AX <sub>2</sub> GGAUA	****	CUXCXAA	GU	GG	UUCGAAU
Asparagine	GUCGGGA	GCUC	AGUUGGUA	GUCGG	CUGUAAA	UGGUC	GUAGG	UUCGAAU
Aspartate	GGGAUUG	GUUC	AAUUGGUA	CCGCC	CUGUCAA	AAGCU	GCGGG	UUCGAGC
Cysteine	GGCGACA	GCC	GAGCGGUA	GGGGA	CUGCAAA	UAUUC	CCCAG	UUCAAAU
Glutamate	GX <sub>2</sub> CX <sub>3</sub>	GX <sub>3</sub>	AGXGGUX <sub>1-3</sub>	CX <sub>2</sub> CX	CUUUCAX	X <sub>2</sub> GX <sub>1-2</sub>	X <sub>3</sub> GX	UUCXAXU
Glutamine	UGGGGCG	GCC	AAGUGGUA	CGGG	UUUUGGU	CUAUGC	GGAGG	UUCGAAU
Glycine	GCGGAUA	GU	CGAAUGGUA	UCUCU	UUGCCAA	AGAC	GCGGG	UUCGAUU
Histidine	GCGGAUG	GCC	AAGUGGAUCAA	GUGGA	UUGUGAA	CAUGC	GCGGG	UUCAAUU
Isoleucine	GCAUCCA	GCU	GAAUGGUAAA	CCCAA	CUCAAAA	AAUUC	GUAGG	UUCAAUU
Leucine	GX <sub>6</sub>	GXG	AAAUXGX <sub>3-4</sub> A	X <sub>3</sub> GX	CUX <sub>4</sub> A	XGX <sub>9-12</sub>	X <sub>3</sub> GG	UUCXAGU
Lysine	GGGUUGC	ACUC	AACGGUA	UCGG	CUUUUAA	CUAGUU	CCGGG	UUCGAGU
Methionine	XCX <sub>5-6</sub>	X <sub>3</sub> G	AGUX <sub>5-6</sub>	****	XUCAUAX	X <sub>2</sub> GUC	AUXGG	UUCAAAU
Phenylalanine	GUCGGGA	GCUC	AGUUGGUA	GAGGA	CUGAAAA	GUGUC	ACCAG	UUCAAAU
Proline	AGGGAUG	GCGC	AGCUUGGUA	UUUGU	UUUGGGU	AUGUC	ACGGG	UUCAAAU
Serine	GGAGAGA	GXC <sub>1-2</sub>	X <sub>4</sub> GX <sub>3-4</sub> A	X <sub>2</sub> GX <sub>1-2</sub>	XUXGXAX	X <sub>4</sub> GX <sub>15-19</sub>	GAGGG	UUCGAAU
Threonine	XGCCX <sub>0-4</sub>	XCUC	AGXGGUA	XCGCX	X <sub>3</sub> GUAA	X <sub>2</sub> GUC	AUCGG	UUCX <sub>3</sub> U
Tryptophan	GCGCUCU	GUUC	AGUUCGGUA	UGGGU	CUCCAAA	AUGUC	GUAGG	UUCAAAU
Tyrosine	GGGUCGA	CCCG	AGCGGUAAA	ACGGA	CUGUAAA	GGCA	GUAGG	UUCAAAU
Valine	AGGGAUA	ACUC	AGCGGUA	UCACC	UUGACGU	AAGUC	AUCAG	UUCGAGC

Note: The asterisk (\*\*\*\*) shows the absence of conserved nucleotide consensus sequence in the respective region of the chloroplast tRNAs.

### 3.2. Diversification of tRNA Structure

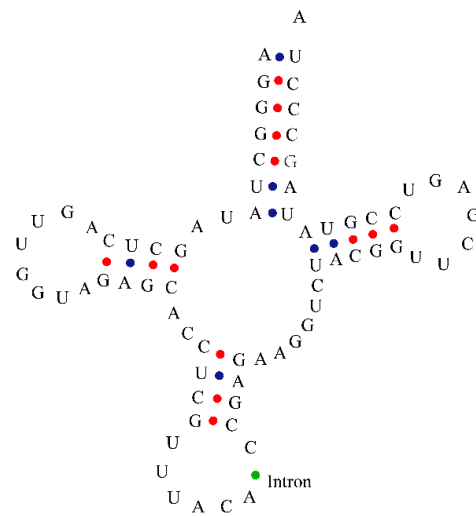
The tRNA feature of possessing various arms and loops was responsible for protein translation. The results of structure analysis showed that the acceptor arm in cotton chloroplast tRNA contained 6 to 7 nt. Among the 369 investigated tRNA sequences, only nine were found to contain 6 nt, while the remaining 360 tRNAs (97.56%) contained 7 nt (Table S3). In addition, the D-arm was observed to contain 2 to 4 nt, among which 20 contained only 2 nt, and 110 (29.81%) contained 3 nt. The remaining tRNAs (64.50%) were observed to possess 4 nt in the D-arm. Moreover, the D-loop contained 7 to 11 nt. For all the involved tRNAs, 84 contained 7 nt in their D-loops; 69 (18.70%) contained 8; 105 (28.46%) contained 9; 40 (10.84%) contained 10; and the rest (18.98%) contained 11 nt (Table S3).

In all 369 tRNAs, the anti-codon arm had mainly 4 to 5 nt. Among them, 319 (86.45%) had 5 nt, and 50 (13.55%) had 4 nt. In addition, 359 tRNAs (97.29%) contained 7 nt in their anti-codon loops and 10 (only 2.71%) contained 9 nt. This showed that the conservative sequence of anti-codon loops was rather typical (Table 3 and Table S3).

For variable loops, 10 tRNAs (2.71%) contained 2; 30 tRNAs (8.13%) contained 3, 67 tRNAs (18.16%) contained 4; 213 tRNAs (57.72%) contained 5; 39 tRNAs (10.57%) contained 6; and 10 tRNAs (2.71%) contained 8 nt (Table S3). While the Ψ-arms of all the analyzed chloroplast tRNAs were observed to contain 5 nt, most tRNAs (349, 94.58%) had 7 nt in their Ψ-loops, apart from several tRNA<sup>Arg</sup> (Table S3).

### 3.3. Chloroplast tRNA Contained Introns

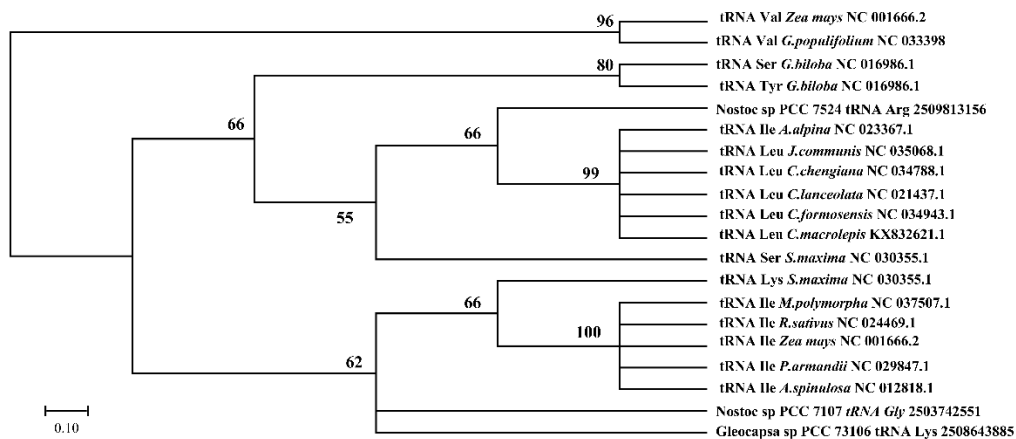
The cotton chloroplast tRNAs had intron annotations according to previous studies. The tRNA<sup>Val</sup> of *G. populifolium* was observed to contain an intron in its anti-codon loop region (Figure 1). The introns in bacterial and plant chloroplast tRNAs had conserved G-A-T-T and C-T-T-C-A consensus sequences (Figure 2). Phylogenetic analysis showed that chloroplast tRNA introns grouped with the introns in cyanobacteria (Figure 3). Introns contained in the same tRNAs of most plants (tRNA<sup>Ile</sup> and tRNA<sup>Leu</sup>) tended to appear in the same branch. This revealed their close phylogenetic relationship. The introns in chloroplast tRNA<sup>Val</sup> of *G. populifolium* and *Zea mays* L. aggregated to the same branch (Figure 3), showing that the introns of corn and cotton have a close phylogenetic relationship.



**Figure 1.** Presence of intron in chloroplast tRNA, found to be located in the tRNA anti-codon loop. The blue dot represents dihydrogen bonds; the red dot represents triple hydrogen bond; and the green dot indicates the position of intron.

	1                    10                    20                    30                    35
	-----+-----+-----+-----+-----
chloroplast tRNA Val_ <i>G. Populifolium</i>	CAATCAGTTCGATCTGTTTTA
chloroplast tRNA Val_ <i>Zea mays</i>	CTGTTTGATCTGTTTTA
chloroplast tRNA Ile_ <i>M. polymorpha</i>	TAGGTGCGATGATTTACTTCA
chloroplast tRNA Ile_ <i>Zea mays</i>	AAGGGGCGATGATTTACTTCA
chloroplast tRNA Ile_ <i>P. armandii</i>	CAGGTGCGAGGATTTACTTCA
chloroplast tRNA Ile_ <i>R. sativus</i>	GGTGCGATGATTTACTTCA
chloroplast tRNA Ile_ <i>A. spinulosa</i>	GGTGCGATTATTTCACTTCA
Nostoc	AACCGT-TCGCGTAG---CGTGGC
Gleocapsa	TTTGCCTCCTTGTCTATTAAGACCGAAGA
chloroplast tRNA Lys_ <i>S. maxima</i>	ATGCCGACTTTCA
chloroplast tRNA Ser_ <i>S. maxima</i>	AATCCGTTGTACG
Consensus	GATTTCTTCA

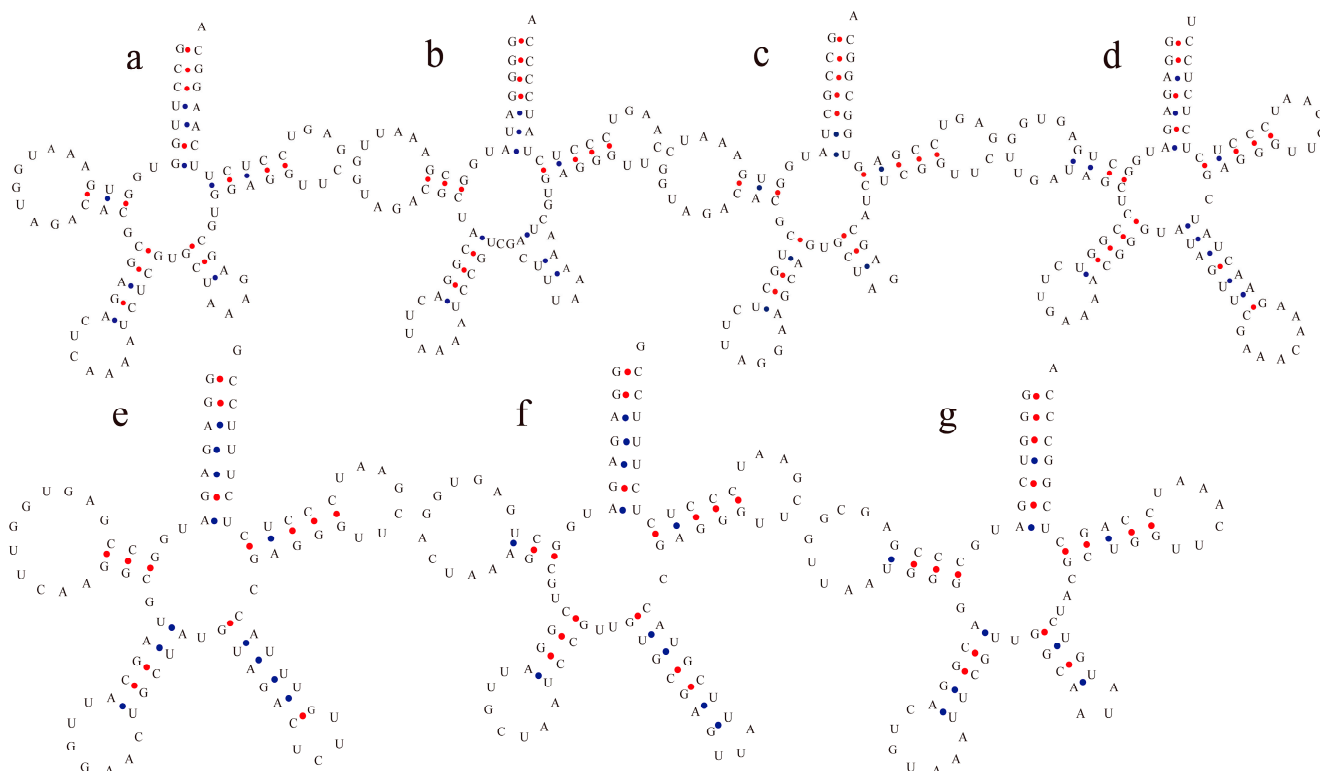
**Figure 2.** Multiple sequence alignment of introns in tRNA of plant chloroplast and cyanobacteria. Introns in bacterial and chloroplast tRNAs had conserved G-A-T-T-T and C-T-T-C-A consensus sequences.



**Figure 3.** Phylogenetic relationship of introns in chloroplast tRNAs (from angiosperms, gymnosperms, ferns, bryophytes, and algae). Introns of chloroplast tRNA grouped with that of cyanobacteria illustrated a common cyanobacterial origin of introns in chloroplasts. Introns in chloroplast tRNA<sup>Val</sup> of *Gossypium populifolium* and *Zea mays* indicated the same evolutionary ancestors of corn and cotton introns. Introns present in the same tRNA of plants (tRNA<sup>Ile</sup> and tRNA<sup>Leu</sup>) tended to appear in the same branch, showing their close phylogenetic relationship.

### 3.4. Chloroplast tRNAs with Non-Typical Features

A few unconventional tRNAs were observed in cotton cp genomes. tRNA<sup>Leu</sup>, tRNA<sup>Ser</sup>, and tRNA<sup>Tyr</sup> were observed to have a loop in the variable region (Figure 4). In these non-typical tRNAs, the anti-codon loop harbored 7 nt with the X-U-X3-A-A consensus sequence, and the stem of the anti-codon loop contained 4 to 5 nt. The variable loop region was observed to have 2 to 8 nt for tRNA<sup>Leu</sup>, tRNA<sup>Ser</sup>, and tRNA<sup>Tyr</sup>. The stem of the variable region contained 3 to 7 nt pairs. Obviously, tRNA<sup>Ser</sup> had mainly 6 or 7 nt pairs (Figure 4). These loop structures in variable regions of tRNAs put forward the question of whether these loops play an important role in the process of protein translation of the chloroplast.



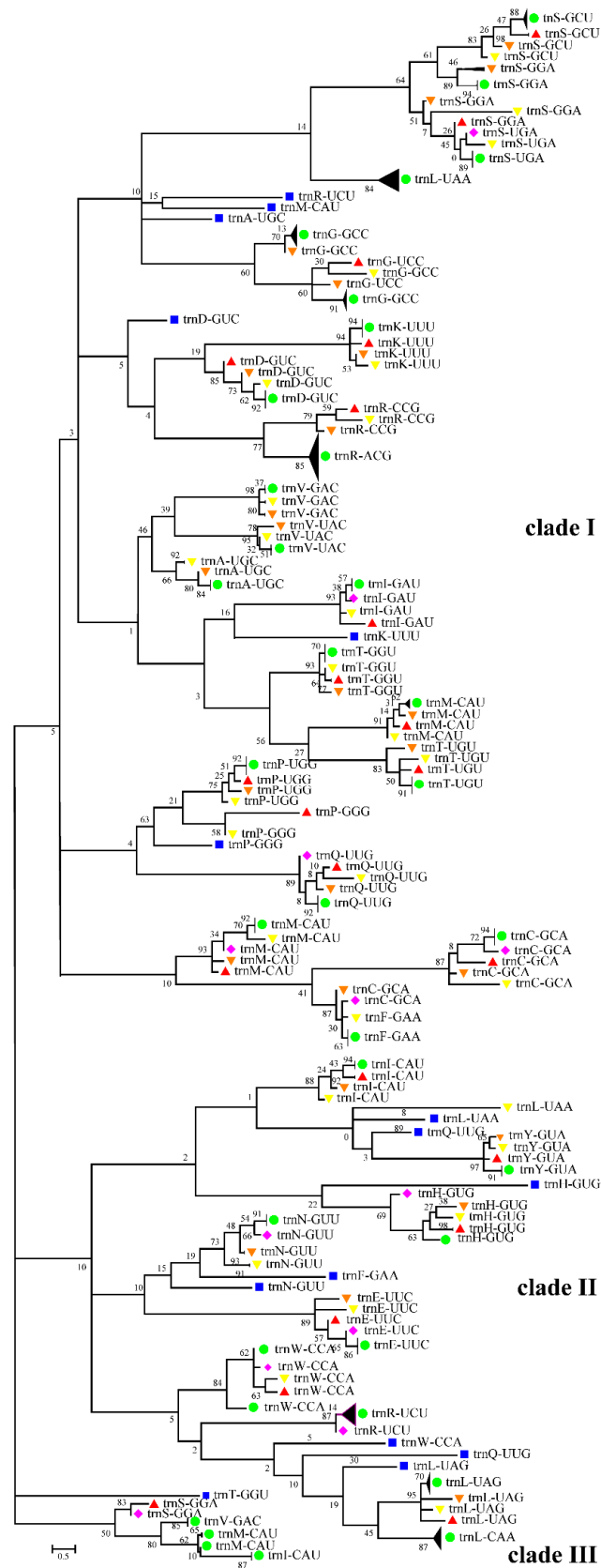
**Figure 4.** Structures of the chloroplast tRNAs showed the presence of a loop structure in the variable region: (a–c) tRNA<sup>Leu</sup>; (d–f) tRNA<sup>Ser</sup>; (g) tRNA<sup>Tyr</sup>. Anti-codon loop had seven nucleotides with the conservative X-U-X3-A-A consensus sequence in these tRNAs. Loop structure of the variable region was observed to contain three to eight nucleotides from tRNA<sup>Leu</sup>, tRNA<sup>Ser</sup>, and tRNA<sup>Tyr</sup>.

### 3.5. Cotton Chloroplast tRNAs Were Derived from Several Evolutionary Ancestors

The phylogenetic tree of cotton and other various species' tRNA sequences from a wide range of taxonomic positions, including algae (*Nostoc* sp. PCC 7524), bryophytes (*Dumortiera hirsuta*), ferns (*Psilotum nudum*), gymnosperms (*Pinus taeda*), and angiosperms (*Arabidopsis lyrata*), presented three major clusters (Table 4). In all, there were 87 groups in integrated clade I, 43 groups in integrated clade II, and 7 groups in integrated clade III. In the phylogenetic tree, tRNA<sup>Ser</sup>, tRNA<sup>Leu</sup>, tRNA<sup>Arg</sup>, tRNA<sup>Val</sup>, tRNA<sup>Ile</sup>, tRNA<sup>Met</sup>, and tRNA<sup>Gln</sup> were polyphyletic, positioned in more than two integrated clades. In integrated clade I, most of tRNA<sup>Met</sup> was clustered into two polyphyletic sub-clades (containing nine groups), one embedded in the tRNA<sup>Thr</sup> group and another adjacent to tRNA<sup>Cys</sup> and tRNA<sup>Phe</sup>, while tRNA<sup>Met</sup> of cyanobacteria was independently clustered to its tRNA<sup>Arg</sup>. In integrated clade II, tRNA<sup>Leu</sup> was closely related to cyanobacterial tRNA<sup>Gln</sup>, even with its polyphyly. tRNA<sup>Trp</sup>, except the cyanobacterial one, was clustered into a monophyletic group, while the cyanobacterial one was embedded into the sister clade of the tRNA<sup>Trp</sup> group. tRNA<sup>Ile</sup> was found in all three clades. Interestingly, tRNA<sup>Leu</sup>, tRNA<sup>Ile</sup>, tRNA<sup>Gln</sup>,



tRNA<sup>Phe</sup>, and tRNA<sup>Arg</sup> present in integrated clade I were also found in clade II, and the tRNAs in integrated clade III were also found in clade I (Figure 5).



**Figure 5.** Phylogenetic tree of chloroplast tRNAs. A, alanine (Ala); R, arginine (Arg); N, asparagine (Asn);

D, aspartate (Asp); C, cysteine (Cys); Q, glutamine (Gln); E, glutamate (Glu); G, glycine (Gly); H, histidine (His); I, isoleucine (Ile); L, leucine (Leu); K, lysine (Lys); M, methionine (Met); F, phenylalanine (Phe); P, proline (Pro); S, serine (Ser); T, threonine (Thr); W, tryptophan (Trp); Y, tyrosine (Tyr); V, valine (Val). The green solid circle represents *Gossypium*; the red solid triangle represents *P. taeda*; the pink solid diamond represents *A. lyrata*; the yellow solid triangle represents *P. nudum*; the orange solid triangle represents *D. hirsuta*; and the blue solid square represents *Nostoc* sp. PCC 7524. Phylogenetic analysis showed the polyphyletic origin of chloroplast tRNAs.

**Table 4.** The distribution of types of tRNAs in the three major clusters of the phylogenetic tree.

Integrated Clades	Types of tRNAs
clade I	tRNA <sup>Ser</sup> , tRNA <sup>Leu</sup> , tRNA <sup>Arg</sup> , tRNA <sup>Met</sup> , tRNA <sup>Ala</sup> , tRNA <sup>Gly</sup> , tRNA <sup>Asp</sup> , tRNA <sup>Lys</sup> , tRNA <sup>Val</sup> , tRNA <sup>Ile</sup> , tRNA <sup>Thr</sup> , tRNA <sup>Pro</sup> , tRNA <sup>Gln</sup> , tRNA <sup>Cys</sup> , tRNA <sup>Phe</sup>
clade II	tRNA <sup>Ile</sup> , tRNA <sup>Leu</sup> , tRNA <sup>Gln</sup> , tRNA <sup>Tyr</sup> , tRNA <sup>His</sup> , tRNA <sup>Asn</sup> , tRNA <sup>Phe</sup> , tRNA <sup>Glu</sup> , tRNA <sup>Trp</sup> , tRNA <sup>Arg</sup>
clade III	tRNA <sup>Thr</sup> , tRNA <sup>Ser</sup> , tRNA <sup>Val</sup> , tRNA <sup>Met</sup> , tRNA <sup>Ile</sup>

### 3.6. Transition/Transversion of tRNAs

tRNAs have evolved with almost equal transition and transversion rates in spite of the small probability of transition or transversion events in tRNAs. In the present study, we found some intriguing phenomena in the substitution rate of cotton chloroplast tRNAs. A transition rate of 16.66 and transversion rate of 16.68 were found in tRNA<sup>Ala</sup>, tRNA<sup>Phe</sup>, tRNA<sup>Asp</sup>, tRNA<sup>Pro</sup>, tRNA<sup>Tyr</sup>, tRNA<sup>Val</sup>, and tRNA<sup>Ile</sup>. This showed that the transversion rate was slightly higher than the transition rate, and these tRNAs evolved with almost the same transition and transversion rates (Figure 6, Table S4). The highest transition rate was 50.00 for tRNA<sup>Trp</sup> and the highest transversion rate was 25.00 for tRNA<sup>His</sup>. Correspondingly, the lowest transition rate (0.00) was observed for tRNA<sup>His</sup>, and the lowest transversion rate (0.00) for tRNA<sup>Trp</sup> (Figure 6). This indicated that the tRNA<sup>Trp</sup> of the cotton cp genome had experienced a high transition rate without any transversion. Similarly, tRNA<sup>His</sup> had experienced a high rate of transversion but no transition. For tRNA<sup>Lys</sup>, tRNA<sup>Arg</sup>, tRNA<sup>Met</sup>, tRNA<sup>Asn</sup>, tRNA<sup>Cys</sup>, tRNA<sup>Ser</sup>, tRNA<sup>Glu</sup>, tRNA<sup>Gly</sup>, and tRNA<sup>Leu</sup>, the transition rate was higher than the transversion rate (Figure 6A). Additionally, the transition rate was two times higher than transversion for tRNA<sup>Arg</sup>, tRNA<sup>Cys</sup>, and tRNA<sup>Glu</sup>, which showed that the evolution of chloroplast tRNA<sup>Arg</sup>, tRNA<sup>Cys</sup>, and tRNA<sup>Glu</sup> tended toward transition rather than transversion. For tRNA<sup>Gln</sup> and tRNA<sup>Thr</sup>, the transversion rate was apparently higher than the transition rate (Figure 6B). This indicated that these tRNA iso-acceptors experienced transversion substitutions more easily than transition. The substitution rates of overall cp tRNAs showed that the average transition rate (23.64) was greater than the transversion rate (13.18) (Figure 6B).

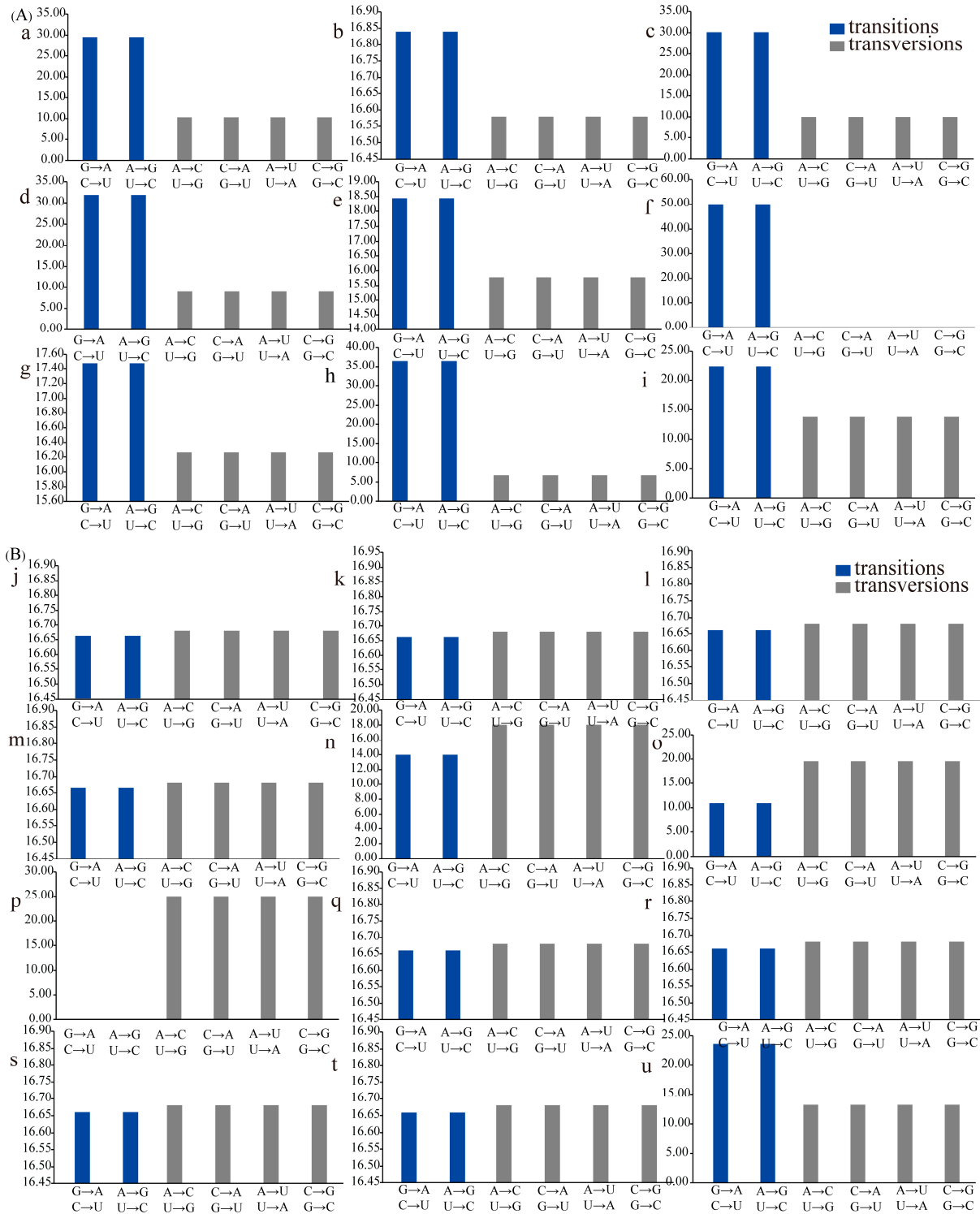
### 3.7. Evolutionary Characteristics of Single Nucleotide Polymorphisms

The biallelic and parallel mutation SNPs of the eight diploid *Gossypium* cp genomes were calculated. There were 2709 SNPs in *Gossypium* cp genomes (Table 5). They were subdivided into coding, intron, and intergenic spacer regions for further analyses. Among the 2709 SNPs, 906 were in coding regions, 299 were in intron regions, and 1504 were in intergenic spacers. The percentage of SNPs to total length was 1.14, 1.39, and 2.90%, respectively. In coding regions, the overall ratio of nonsynonymous mutations to synonymous mutations (dN/dS) was 3.04.

### 3.8. Mutation Rate of Chloroplast Genome

The divergence time of *Gossypium* was speculated to be 9.8 Mya from *Theobroma* [52,53]. The total length of the eight diploid cotton cp genomes was 156,796 bp. The mutation rate ( $\mu_1$ ) of the protein-coding genes and the mutation rate ( $\mu_2$ ) of the tRNA genes were calculated using the length of genomes, the number of indel mutations (522 for protein-coding

genes and 133 for tRNA genes), and the divergent time. The value of  $\mu_1$  of  $0.34 \times 10^{-9}$  per site per year for protein-coding genes and the value of  $\mu_2$  of  $0.09 \times 10^{-9}$  per site per year for tRNA genes were obtained, respectively.



**Figure 6.** Rates of transition (blue) and transversion (gray) of chloroplast tRNAs. (A) a–i refer to R, Arg; S, Ser; C, Cys; E, Glu; G, Gly; W, Trp; L, Leu; K, Lys; M, Met. (B) j–u refer to A, Ala; P, Pro; N Asn; D, Asp; T, Thr; Q, Gln; H, His; I, Ile; Y, Tyr; V, Val; F, Phe; overall.

**Table 5.** Genomic distribution of biallelic single nucleotide polymorphic loci in the eight chloroplast genomes of diploid *Gossypium* plants.

Genome Region	Length (bp)	Value	%
Total substitutions	162,231	2709	1.67
Coding regions	79,244	906	1.14
Non-synonymous	/	681.58	0.86
Synonymous	/	224.42	0.28
dN/dS	/	3.04	/
Intron	21,443	299	1.39
Intergenic spacer	51,920	1504	2.90

### 3.9. tRNA Duplication/Loss Events

Besides substitution, duplication and loss events of genes had a vital influence on their evolution. The analysis of duplication or loss events indicated that the investigated cotton chloroplast tRNA genes had experienced 226 duplication and 93 co-duplication events (Figure S1). However, only 63 loss events were observed (Figure S1, Table S5). The results showed that the duplication of genes was greater than the loss of genes for all of the tRNAs.

## 4. Discussion

The nucleotide composition of tRNA is closely related to its senior structure, which is responsible for the translation process. In many species, the tRNA family is conserved in evolution [56,57]. As one of the major gene components of semi-autonomous chloroplast, tRNAs were shown to have several basic conserved genomic features. tRNA<sup>Leu</sup> and tRNA<sup>Ser</sup> were observed to have 80 nt or more. These two tRNA isotypes were also found to harbor more than 83 nt in Adoxaceae plants, which shows that the gene sequences of tRNA<sup>Ser</sup> and tRNA<sup>Leu</sup> are longer than other tRNAs [58]. In addition, some tRNA isoacceptors were not observed in the cp genome of *Gossypium*, which was similar to that of the species in Gramineae [28]. Additionally, this is perhaps related to the codon usage and wobble base pairing in the genomic constitution of plants [59]. Furthermore, selenocysteine and suppressor tRNAs were found lacking in the cotton cp genome, which were present in *Oryza sativa*, *Sachharum officinarum* L., *Sorghum bicolor* (L.) Moench, *Triticum aestivum* L., and *Zea mays*. This may be related to the biotoxin accumulation ability of these two tRNAs [29]. The differences between the quantity (36 or 37) and anti-codon types (28 or 29) of tRNAs in diploid and allotetraploid cotton species were not significant. This may be related to the stability and conservation of tRNA genes and the long cultivation of cotton species, which reduced the level of interspecific diversity [60,61].

The tRNA family was conserved in its genomic composition. In this study, most of the involved tRNAs were observed to harbor the conservative sequence U-U-C-X3-U at their Ψ-loop (Table 3). Additionally, in Adoxaceae, U-U-C-A conservative nucleotides at its Ψ-loop region were found [58]. This showed that the short consensus sequences might be important components of the molecular recognizer in the Ψ-loop, and are associated with the recognition of ribosomal RNA during the translation process [62]. In our study, we found that the acceptor arm of cotton chloroplast tRNA contained 6 to 7 nt; the D-arm contained 2 to 4 nt; the D-loop contained 7 to 11 nt; the anti-codon loops contained 7 and 9 nt; the Ψ-arms contained 5 nt; and the Ψ-loops had 7 nt. Additionally, in gymnosperm chloroplast genomes, the acceptor arm of tRNAs harbors 6 bp to 7 bp; the D-arm has 3 bp or 4 bp; and the D-loop contains 7 nt to 11 nt; the anti-codon loop contains 7 nt; the Ψ-arm contains 5 bp; and the Ψ-loop has 7 nt [63]. Our results and previous findings together suggest that chloroplast RNAs are significantly conserved, though a few tRNAs contain rare secondary structures [64,65]. In the present study, some tRNA<sup>Leu</sup>, tRNA<sup>Ser</sup>, and tRNA<sup>Tyr</sup> were found in the variable region including a loop structure. The presence of this unconventional structure in the variable region exhibited the structural variation

existing in tRNAs and this might be relevant to the maintenance of tRNA structures and the interaction with the D-arm and  $\Psi$ -arm [66].

Introns are previously reported in archaeal and eukaryotic genomes that break the continuity of numerous eukaryotic genes [67]. Here, the intron was observed in chloroplast tRNAs of *G. populifolium*. This is consistent with previous reports in other organisms [68,69]. In general, the group I intron has a few hundred nucleotides because of its built-in ribozyme unit [70]. tRNAs contain sequences of less than 100 polynucleotides that fold into a clover-type secondary structure [71]. Thus, the intron we observed in this study might be a part of intron I that was contained within the cotton chloroplast genome.

Phylogenetic analysis revealed that the introns in chloroplast tRNA<sup>Val</sup> of *Gossypium populifolium* and *Zea mays* aggregated to the same branch (Figure 3), suggesting that the introns of corn and cotton have a close phylogenetic relationship. Additionally, chloroplast tRNAs with introns were grouped with cyanobacteria, providing supportive evidence for a common cyanobacterial lineage source of cp tRNAs [58].

In addition to the anti-codon CAU of tRNA<sup>Met</sup>, tRNA<sup>Ile</sup> was observed to code the anti-codon CAU in the cotton chloroplast genome. This may be associated with the modification of tRNA [72]. Three types of tRNAs—tRNA<sup>fMet</sup>, tRNA<sup>Met</sup>, and tRNA<sup>klle</sup>, with anti-codon CAU—were found in a bacterial genome, and tRNA<sup>klle</sup> was able to identify the codon of isoleucine instead of methionine after anti-codon modification [73]. For the tRNA<sup>Ile</sup> CAU observed in the genomes of plants, the first position in the anti-codon of these tRNAs has a lysidine-like nucleotide. On the other hand, a methionine-discerning anti-codon CAT was found in the genes. After the modification of the C residue of the CAT anti-codon, the mature tRNA shows isoleucine-identifying activity rather than methionine-identifying activity [74,75].

The presence of three obvious clusters and diverse groupings, and the appearance of tRNA groups from different plants with a wide range of taxonomic positions in integrated clade I and II, clade I and III, and clade I, II, and III of the phylogenetic tree, indicate frequent duplication and divergence during their evolution. In addition, tRNA<sup>Ser</sup>, tRNA<sup>Leu</sup>, tRNA<sup>Arg</sup>, tRNA<sup>Val</sup>, tRNA<sup>Ile</sup>, tRNA<sup>Met</sup>, and tRNA<sup>Gln</sup> were found in more than two integrated clades. This indicates their multiple evolutionary origins. Diverse groupings and the overlapping of tRNA groups from different plants suggest that the tRNAs have several inferred ancestors, including tRNA<sup>Met</sup>, tRNA<sup>Ile</sup>, etc., in the evolutionary history [63]. The interlaced emergence of tRNA groups (tRNA<sup>Met</sup>, tRNA<sup>Thr</sup>, tRNA<sup>Arg</sup>, tRNA<sup>Trp</sup>, tRNA<sup>Val</sup>, and tRNA<sup>Ser</sup>) also suggests that evolutionary relationships among these tRNA types are relatively close.

Just as the cases with the most genetic variations, the transition rate was greater than the transversion rate of *Gossypium* cp tRNAs at the overall level. However, tRNA<sup>Trp</sup> and tRNA<sup>His</sup> showed different substitutive choices, which may be caused by various factors, such as their neighbor bases and the efficiency of the repair system of DNA strands [76]. The percentage of SNPs in coding regions, intron regions, and intergenic spacers compared to the total cotton cp genomes was 1.14, 1.39, and 2.90%, respectively, which implies that the density of SNPs was different in the cotton cp genome and the intergenic spacer of *Gossypium* was more variable than intron regions. This may be associated with the richness of A/T and G/C repetitive units in certain regions of the cp genome. It also implies that the evolutionary mutation potential of different positions in the genome is unbalanced [77]. In coding regions, the overall ratio of non-synonymous to synonymous mutations (dN/dS) was 3.04, showing that non-synonymous mutations had been fixed in the cotton cp genome. In the involved diploid *Gossypium* cp genomes, the mutation rate ( $\mu_1$ ),  $0.34 \times 10^{-9}$  per site per year of protein-coding genes, was significantly higher than the mutation rate ( $\mu_2$ ),  $0.09 \times 10^{-9}$  per site per year of tRNA genes (more than three times than that of  $\mu_2$ ), which also implies that compared with protein-coding genes, tRNA genes and structures have considerable evolutionary stability [78]. In addition, it should be mentioned that the genomic data employed in our study still have limitations. Perhaps the addition of more diverse gene sequences would provide more interesting results.



In concert with multiple factors of tRNA evolution, most important gene functions have evolved from gene duplication events [79,80]. In this study, there were about five times as many duplication events as loss events. Many studies have also shown that frequent duplication greatly promotes evolution and functional diversification in gene families [81,82]. This may provide further insights to confirm that the evolution of cotton tRNAs derived from polyphyletic evolutionary ancestors.

## 5. Conclusions

Cotton chloroplast tRNAs encode 28 or 29 anti-codon types, and 36 or 37 anti-codon-specific tRNAs with lengths ranging from 60 (tRNA<sup>Gly</sup>) to 93 nucleotides (tRNA<sup>Ser</sup>). Thirty-three anti-codon types including AGC-tRNA<sup>Ala</sup> are absent in the cp genome of *Gossypium*. The CAU anti-codon is encoded in both tRNA<sup>Met</sup> and tRNA<sup>Ile</sup>. The acceptor arm of cotton chloroplast tRNA contains 6 to 7 nt; the D-arm contains 2 to 4 nt; the D-loop contains 7 to 11 nt; the anti-codon loop contains 7 and 9 nt; the Ψ-arms contains 5 nt; and the Ψ-loop has 7 nt. The Ψ-arm of cotton chloroplast tRNAs contains the G-G consensus sequence, and the Ψ-loop contains the conserved U-U-C-X3-U motifs. Additionally, cotton chloroplast tRNAs were found to contain introns and a few tRNA<sup>Leu</sup>, tRNA<sup>Ser</sup>, and tRNA<sup>Tyr</sup> were observed to have a loop in the variable region. Furthermore, phylogenetic analysis suggests that tRNAs possibly have several inferred ancestors, including tRNA<sup>Met</sup>, tRNA<sup>Ile</sup>, etc., in the evolutionary history. On the other hand, the average transition rate of all involved cp tRNAs was greater than their transversion rate. The density of SNPs was unbalanced and the intergenic spacer of *Gossypium* was more variable than intron regions. Gene duplication events (226 duplication and 93 co-duplication) have occurred more frequently than gene loss events (63) in cotton chloroplast tRNAs. These results provide helpful insights into the detailed characteristics and evolutionary variation of the tRNA family.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/genes12060822/s1>. Figure S1: Duplication and loss events of the chloroplast tRNAs. Blue: duplication events; gray: loss events; D: duplication node; cD: conditional duplication node. Analysis indicates that the chloroplast tRNAs underwent frequent duplication events during their evolution with subsequent diversification, Table S1: Information registration of the analyzed tRNA genes, Table S2: Distribution of anti-codons in cotton chloroplast genome, Table S3: Nucleotide composition in acceptor arm (AA), D-arm (DA), D-loop (DL), anti-codon arm (ACA), anti-codon loop (ACL), variable loop (VL), pseudouridine arm (ΨA), and pseudouridine loop (ΨL) of chloroplast tRNA, Table S4: Results of transition and transversion rates of chloroplast tRNAs, Table S5: Loss events of the chloroplast tRNAs.

**Author Contributions:** Conceptualization, Z.-H.L. and T.-T.Z.; methodology, T.-T.Z., Y.Y., and X.-Y.S.; software, X.-L.Z.; validation, X.-Y.G., J.-J.Z., and X.-Y.S.; formal analysis, K.-H.Z.; investigation, C.-B.Z.; resources, W.L.; data curation, D.-G.Y. and X.-F.M.; writing—review and editing, Z.-H.L., Y.Y., and T.-T.Z.; supervision, Z.-H.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by grants from the National Key R&D Program for Crop Breeding (2016YFD0100306), the National Natural Science Foundation of China (No. 31401431), Shaanxi Science and Technology Innovation Team (2019TD-012), Fourth National Survey of Traditional Chinese Medicine Resources (2018-43 and 2019-39), and the Open Foundation of Key Laboratory of Resource Biology and Biotechnology in Western China (Ministry of Education) (grant nos. ZSK2017007 and ZSK2019008).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Acknowledgments:** We thank Qiu-Yi Zhong for her kindly help for the manuscript improvements.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Arnon, D.I. The chloroplast as a complete photosynthetic unit. *Science* **1955**, *122*, 9–16. [[CrossRef](#)]
2. Roughan, P.G.; Holland, R.; Slack, C.R. On the control of long-chain-fatty acid synthesis in isolated intact spinach (*Spinacia oleracea*) chloroplasts. *Biochem. J.* **1979**, *184*, 193–202. [[CrossRef](#)]
3. Neuhaus, H.E.; Emes, M.J. Nonphotosynthetic metabolism in plastids. *Annu. Rev. Plant Biol.* **2000**, *51*, 111–140. [[CrossRef](#)] [[PubMed](#)]
4. Spetea, C.; Hundal, T.; Lundin, B.; Heddad, M.; Adamska, I.; Andersson, B. Multiple evidence for nucleotide metabolism in the chloroplast thylakoid lumen. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 1409–1414. [[CrossRef](#)] [[PubMed](#)]
5. Holsinger, K.E.; Soltis, P.S.; Soltis, D.E.; Doyle, J.J. Molecular systematics of plants. *Syst. Bot.* **1993**, *18*, 539. [[CrossRef](#)]
6. Gross, J.; Meurer, J.; Bhattacharya, D. Evidence of a chimeric genome in the cyanobacterial ancestor of plastids. *BMC Evol. Biol.* **2008**, *8*, 117. [[CrossRef](#)] [[PubMed](#)]
7. Lin, C.-P.; Wu, C.-S.; Huang, Y.-Y.; Chaw, S.-M. The complete chloroplast genome of ginkgo biloba reveals the mechanism of inverted repeat contraction. *Genome Biol. Evol.* **2012**, *4*, 374–381. [[CrossRef](#)] [[PubMed](#)]
8. Nguyen, P.A.T.; Kim, J.S.; Kim, J.-H. The complete chloroplast genome of colchicine plants (*Colchicum autumnale* L. and *Gloriosa superba* L.) and its application for identifying the genus. *Planta* **2015**, *242*, 223–237. [[CrossRef](#)] [[PubMed](#)]
9. Ruf, S.; Karcher, D.; Bock, R. Determining the transgene containment level provided by chloroplast transformation. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 6998–7002. [[CrossRef](#)] [[PubMed](#)]
10. Kuo, L.-Y.; Tang, T.-Y.; Li, F.-W.; Su, H.-J.; Chiou, W.-L.; Huang, Y.-M.; Wang, C.-N. Organelle genome inheritance in *Deparia* ferns (Athryiaceae, Aspleniineae, Polypodiales). *Front. Plant Sci.* **2018**, *9*, 486. [[CrossRef](#)]
11. Wang, R.-J.; Cheng, C.-L.; Chang, C.-C.; Wu, C.-L.; Su, T.-M.; Chaw, S.-M. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evol. Biol.* **2008**, *8*, 36. [[CrossRef](#)] [[PubMed](#)]
12. Fish, L.E.; Jagendorf, A.T. High rates of protein synthesis by isolated chloroplasts. *Plant Physiol.* **1982**, *70*, 1107–1114. [[CrossRef](#)]
13. Shinozaki, K.; Hayashida, N.; Sugiura, M. *Nicotiana* chloroplast genes for components of the photosynthetic apparatus. *Photosynth. Res.* **1988**, *18*, 7–31. [[CrossRef](#)] [[PubMed](#)]
14. Kirchner, S.; Ignatova, Z. Emerging roles of tRNA in adaptive translation, signalling dynamics and disease. *Nat. Rev. Genet.* **2014**, *16*, 98–112. [[CrossRef](#)]
15. Brumfield, R.T.; Beerli, P.; Nickerson, D.A.; Edwards, S.V. Single nucleotide polymorphisms (SNPs) as markers in phylogeography. *Trends Ecol. Evol.* **2003**, *18*, 249–256. [[CrossRef](#)]
16. Rafalski, A. Applications of single nucleotide polymorphisms in crop genetics. *Curr. Opin. Plant Biol.* **2002**, *5*, 94–100. [[CrossRef](#)]
17. Holley, R.W.; Apgar, J.; Everett, G.A.; Madison, J.T.; Marquisee, M.; Merrill, S.H.; Penswick, J.R.; Zamir, A. Structure of a ribonucleic acid. *Science* **1965**, *147*, 1462–1465. [[CrossRef](#)] [[PubMed](#)]
18. Nicoghosian, K.; Bigras, M.; Sankoff, D.; Cedergren, R. Archetypical features in tRNA families. *J. Mol. Evol.* **1987**, *26*, 341–346. [[CrossRef](#)]
19. Giegé, R.; Puglisi, J.D.; Florentz, C. tRNA structure and aminoacylation efficiency. *Prog. Nucleic Acid Res. Mol. Biol.* **1993**, *45*, 129–206. [[CrossRef](#)] [[PubMed](#)]
20. De Pouplana, L.R.; Dedon, P.C. More than an adaptor molecule: The emerging role of tRNA in cell signaling and disease. *FEBS Lett.* **2014**, *588*, 4267. [[CrossRef](#)] [[PubMed](#)]
21. Blee, E.; Joyard, J. Envelope membranes from spinach chloroplasts are a site of metabolism of fatty acid hydroperoxides. *Plant Physiol.* **1996**, *110*, 445–454. [[CrossRef](#)]
22. Kanai, A. Disrupted tRNA genes and tRNA fragments: A perspective on tRNA gene evolution. *Life* **2015**, *5*, 321–331. [[CrossRef](#)]
23. Anderson, P.; Ivanov, P. tRNA fragments in human health and disease. *FEBS Lett.* **2014**, *588*, 4297–4304. [[CrossRef](#)]
24. Kashdan, M.A.; Dudock, B.S. The gene for a spinach chloroplast isoleucine tRNA has a methionine anticodon. *J. Biol. Chem.* **1982**, *257*, 11191–11194. [[CrossRef](#)]
25. Yasukawa, T.; Kirino, Y.; Ishii, N.; Holt, I.; Jacobs, H.T.; Makifuchi, T.; Fukuhara, N.; Ohta, S.; Suzuki, T.; Watanabe, K. Wobble modification deficiency in mutant tRNAs in patients with mitochondrial diseases. *FEBS Lett.* **2005**, *579*, 2948–2952. [[CrossRef](#)] [[PubMed](#)]
26. Morscher, R.J.; Ducker, G.S.; Li, S.H.-J.; Mayer, J.A.; Gitai, Z.; Sperl, W.; Rabinowitz, J.D. Mitochondrial translation requires folate-dependent tRNA methylation. *Nat. Cell Biol.* **2018**, *554*, 128–132. [[CrossRef](#)] [[PubMed](#)]
27. Mohanta, T.K.; Syed, A.S.; Ameen, F.; Bae, H. Novel genomic and evolutionary perspective of cyanobacterial tRNAs. *Front. Genet.* **2017**, *8*, 200. [[CrossRef](#)] [[PubMed](#)]
28. Mohanta, T.K.; Bae, H. Analyses of genomic tRNA reveal presence of novel tRNAs in *Oryza sativa*. *Front. Genet.* **2017**, *8*, 90. [[CrossRef](#)]
29. Mohanta, T.K.; Khan, A.L.; Hashem, A.; Allah, E.F.A.; Yadav, D.; Al-Harrasi, A. Genomic and evolutionary aspects of chloroplast tRNA in monocot plants. *BMC Plant Biol.* **2019**, *19*, 39. [[CrossRef](#)]
30. Fryxell, P.A. A classification of *Gossypium* L. (Malvaceae). *Taxon* **1969**, *18*, 585–591. [[CrossRef](#)]
31. Beasley, J.O. The origin of American Tetraploid *Gossypium* species. *Am. Nat.* **1940**, *74*, 285–286. [[CrossRef](#)]
32. Li, F.; Fan, G.; Lu, C.; Xiao, G.; Zou, C.; Kohel, R.J.; Ma, Z.; Shang, H.; Ma, X.; Wu, J.; et al. Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nat. Biotechnol.* **2015**, *33*, 524–530. [[CrossRef](#)]

33. Zhang, T.; Hu, Y.; Jiang, W.; Fang, L.; Guan, X.; Chen, J.; Zhang, J.; A Sasaki, C.; E Scheffler, B.; Stelly, D.; et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat. Biotechnol.* **2015**, *33*, 531–537. [[CrossRef](#)]
34. Sun, R.; Wang, K.; Guo, T.; Jones, D.C.; Cobb, J.; Zhang, B.; Wang, Q. Genome-wide identification of auxin response factor (ARF) genes and its tissue-specific prominent expression in *Gossypium raimondii*. *Funct. Integr. Genom.* **2015**, *15*, 481–493. [[CrossRef](#)] [[PubMed](#)]
35. Yuan, D.; Tang, Z.; Wang, M.; Gao, W.; Tu, L.; Jin, X.; Chen, L.; He, Y.; Zhang, L.; Zhu, L.; et al. The genome sequence of Sea-Island cotton (*Gossypium barbadense*) provides insights into the allopolyploidization and development of superior spinnable fibres. *Sci. Rep.* **2016**, *5*, 17662. [[CrossRef](#)] [[PubMed](#)]
36. He, Q.; Jones, D.C.; Li, W.; Xie, F.; Ma, J.; Sun, R.; Wang, Q.; Zhu, S.; Zhang, B. Genome-wide identification of R2R3-MYB genes and expression analyses during abiotic stress in *Gossypium raimondii*. *Sci. Rep.* **2016**, *6*, 22980. [[CrossRef](#)] [[PubMed](#)]
37. Ma, L.-F.; Li, Y.; Chen, Y.; Li, X.-B. Improved drought and salt tolerance of *Arabidopsis thaliana* by ectopic expression of a cotton (*Gossypium hirsutum*) CBF gene. *Plant Cell Tissue Organ Cult.* **2016**, *124*, 583–598. [[CrossRef](#)]
38. Salih, H.; Gong, W.; He, S.; Sun, G.; Sun, J.; Du, X. Genome-wide characterization and expression analysis of MYB transcription factors in *Gossypium hirsutum*. *BMC Genet.* **2016**, *17*, 1–12. [[CrossRef](#)]
39. Sun, Q.; Wang, G.; Zhang, X.; Zhang, X.; Qiao, P.; Long, L.; Yuan, Y.; Cai, Y. Genome-wide identification of the TIFY gene family in three cultivated *Gossypium* species and the expression of JAZ genes. *Sci. Rep.* **2017**, *7*, 42418. [[CrossRef](#)]
40. Wu, Y.; Liu, F.; Yang, D.-G.; Li, W.; Zhou, X.-J.; Pei, X.-Y.; Liu, Y.-G.; He, K.-L.; Zhang, W.-S.; Ren, Z.-Y.; et al. Comparative chloroplast genomics of *Gossypium* species: Insights into repeat sequence variations and phylogeny. *Front. Plant Sci.* **2018**, *9*, 376. [[CrossRef](#)]
41. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [[CrossRef](#)] [[PubMed](#)]
42. Lowe, T.M.; Eddy, S.R. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **1997**, *25*, 0955–0964. [[CrossRef](#)] [[PubMed](#)]
43. Lowe, T.M.; Chan, P.P. tRNAscan-SE On-line: Integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* **2016**, *44*, W54–W57. [[CrossRef](#)]
44. Mitchell, C. MultAlin—multiple sequence alignment. *Bioinformatics* **1993**, *9*, 614. [[CrossRef](#)]
45. Mohanta, T.K.; Arora, P.K.; Mohanta, N.; Parida, P.; Bae, H. Identification of new members of the MAPK gene family in plants shows diverse conserved domains and novel activation loop variants. *BMC Genom.* **2015**, *16*, 1–20. [[CrossRef](#)] [[PubMed](#)]
46. Mohanta, T.K.; Mohanta, N.; Parida, P.; Panda, S.K.; Ponpandian, L.N.; Bae, H. Genome-wide identification of mitogen-activated protein kinase gene family across fungal lineage shows presence of novel and diverse activation loop motifs. *PLoS ONE* **2016**, *11*, e0149861. [[CrossRef](#)]
47. Kumar, S.; Nei, M.; Dudley, J.; Tamura, K. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief. Bioinform.* **2008**, *9*, 299–306. [[CrossRef](#)] [[PubMed](#)]
48. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [[CrossRef](#)]
49. Tamura, K. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Mol. Biol. Evol.* **1992**, *9*, 678–687. [[CrossRef](#)]
50. Librado, P.; Rozas, J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **2009**, *25*, 1451–1452. [[CrossRef](#)]
51. Saitou, N.; Ueda, S. Evolutionary rates of insertion and deletion in noncoding nucleotide sequences of primates. *Mol. Biol. Evol.* **1994**, *11*, 504–512. [[CrossRef](#)]
52. Chen, Z.; Feng, K.; Grover, C.E.; Li, P.; Liu, F.; Wang, Y.; Xu, Q.; Shang, M.; Zhou, Z.; Cai, X.; et al. Chloroplast DNA structural variation, phylogeny, and age of divergence among Diploid Cotton species. *PLoS ONE* **2016**, *11*, e0157183. [[CrossRef](#)] [[PubMed](#)]
53. Grover, C.E.; Gallagher, J.P.; Jareczek, J.J.; Page, J.T.; Udall, J.A.; Gore, M.; Wendel, J.F. Re-evaluating the phylogeny of allopolyploid *Gossypium* L. *Mol. Phylogenet. Evol.* **2015**, *92*, 45–52. [[CrossRef](#)] [[PubMed](#)]
54. Chen, K.; Durand, D.; Farach-Colton, M. NOTUNG: A program for dating gene duplications and optimizing gene family trees. *J. Comput. Biol.* **2000**, *7*, 429–447. [[CrossRef](#)]
55. Vernet, B.; Stolzer, M.; Goldman, A.; Durand, D. Reconciliation with non-binary species trees. *J. Comput. Biol.* **2008**, *15*, 981–1006. [[CrossRef](#)] [[PubMed](#)]
56. Kumazawa, Y.; Nishida, M. Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics. *J. Mol. Evol.* **1993**, *37*, 380–398. [[CrossRef](#)]
57. Shepotinovskaya, I.; Uhlenbeck, O.C. tRNA residues evolved to promote translational accuracy. *RNA* **2013**, *19*, 510–516. [[CrossRef](#)]
58. Zhong, Q.; Fu, X.; Zhang, T.; Zhou, T.; Yue, M.; Liu, J.; Li, Z. Phylogeny and evolution of chloroplast tRNAs in Adoxaceae. *Ecol. Evol.* **2021**, *11*, 1294–1309. [[CrossRef](#)]
59. Tong, K.-L.; Wong, J.-F. Anticodon and wobble evolution. *Gene* **2004**, *333*, 169–177. [[CrossRef](#)]
60. Iqbal, M.J.; Aziz, N.; Saeed, N.A.; Zafar, Y.; Malik, K.A. Genetic diversity evaluation of some elite cotton varieties by RAPD analysis. *Theor. Appl. Genet.* **1997**, *94*, 139–144. [[CrossRef](#)]

61. Kumar, P.; Anaya, J.; Mudunuri, S.B.; Dutta, A. Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets. *BMC Biol.* **2014**, *12*, 1–14. [[CrossRef](#)]
62. Charette, M.W.G.M. Pseudouridine in RNA: What, where, how, and why. *IUBMB Life* **2000**, *49*, 341–351. [[CrossRef](#)] [[PubMed](#)]
63. Zhang, T.-T.; Hou, Y.-K.; Yang, T.; Zhang, S.-Y.; Yue, M.; Liu, J.; Li, Z. Evolutionary analysis of chloroplast tRNA of Gymnosperm revealed the novel structural variation and evolutionary aspect. *PeerJ* **2020**, *8*, e10312. [[CrossRef](#)]
64. Nardi, F.; Carapelli, A.; Fanciulli, P.P.; Dallai, R.; Frati, F. The complete mitochondrial DNA sequence of the basal Hexapod *Tetradontophora bielensis*: Evidence for Heteroplasm and tRNA translocations. *Mol. Biol. Evol.* **2001**, *18*, 1293–1304. [[CrossRef](#)] [[PubMed](#)]
65. Jühling, T.; Duchardt-Ferner, E.; Bonin, S.; Wöhnert, J.; Pütz, J.; Florentz, C.; Betat, H.; Sauter, C.; Mörl, M. Small but large enough: Structural properties of armless mitochondrial tRNAs from the nematode *Romanomermis culicivorax*. *Nucleic Acids Res.* **2018**, *46*, 9170–9180. [[CrossRef](#)]
66. Wang, W.; Chen, X.; Wolin, S.L.; Xiong, Y. Structural basis for tRNA mimicry by a bacterial Y RNA. *Structure* **2018**, *26*, 1635–1644. [[CrossRef](#)]
67. Yoshihisa, T. Handling tRNA introns, archaeal way and eukaryotic way. *Front. Genet.* **2014**, *5*, 213. [[CrossRef](#)]
68. Kawach, O.; Voß, C.; Wolff, J.; Hadfi, K.; Maier, U.-G.; Zauner, S. Unique tRNA introns of an enslaved algal cell. *Mol. Biol. Evol.* **2005**, *22*, 1694–1701. [[CrossRef](#)] [[PubMed](#)]
69. Hoser, S.M.; Hoffmann, A.; Meindl, A.; Gamper, M.; Fallmann, J.; Bernhart, S.H.; Müller, L.; Ploner, M.; Misslinger, M.; Kremser, L.; et al. Intronic tRNAs of mitochondrial origin regulate constitutive and alternative splicing. *Genome Biol.* **2020**, *21*, 1–35. [[CrossRef](#)]
70. Adams, P.L.; Stahley, M.R.; Kosek, A.B.; Wang, J.; Strobel, S.A. Crystal structure of a self-splicing group I intron with both exons. *Nat. Cell Biol.* **2004**, *430*, 45–50. [[CrossRef](#)]
71. Wilusz, J.E. Controlling translation via modulation of tRNA levels. *Wiley Interdiscip. Rev. RNA* **2015**, *6*, 453–470. [[CrossRef](#)]
72. Köhrer, C.; Mandal, D.; Gaston, K.W.; Grosjean, H.; Limbach, P.A.; Rajbhandary, U.L. Life without tRNA<sup>Ile</sup>-lysine synthetase: Translation of the isoleucine codon AUA in *Bacillus subtilis* lacking the canonical tRNA<sup>Ile</sup>. *Nucleic Acids Res.* **2013**, *42*, 1904–1915. [[CrossRef](#)] [[PubMed](#)]
73. Sonawane, K.D.; Tewari, R. Conformational Preferences of Hypermodified Nucleoside Lysidine (k2C) Occurring at “Wobble” position in Anticodon Loop of tRNA<sup>Ile</sup>. *Nucleosides Nucleotides Nucleic Acids* **2008**, *27*, 1158–1174. [[CrossRef](#)]
74. Weber, F.; Dietrich, A.; Weil, J.-H.; Maréchal-Drouard, L. A potato mitochondrial isoleucine tRNA is coded for by a mitochondrial gene possessing a methionine anticodon. *Nucleic Acids Res.* **1990**, *18*, 5027–5030. [[CrossRef](#)] [[PubMed](#)]
75. Soma, A.; Ikeuchi, Y.; Kanemasa, S.; Kobayashi, K.; Ogasawara, N.; Ote, T.; Kato, J.-I.; Watanabe, K.; Sekine, Y.; Suzuki, T. An RNA-modifying enzyme that governs both the codon and amino acid specificities of isoleucine tRNA. *Mol. Cell* **2003**, *12*, 689–698. [[CrossRef](#)]
76. Shrestha, B.; Weng, M.-L.; Theriot, E.C.; Gilbert, L.E.; Ruhlman, T.A.; Krosnick, S.E.; Jansen, R.K. Highly accelerated rates of genomic rearrangements and nucleotide substitutions in plastid genomes of *Passiflora* subgenus *Decaloba*. *Mol. Phylogenet. Evol.* **2019**, *138*, 53–64. [[CrossRef](#)]
77. Zhao, Z.; Fu, Y.-X.; Hewett-Emmett, D.; Boerwinkle, E. Investigating single nucleotide polymorphism (SNP) density in the human genome and its implications for molecular evolution. *Gene* **2003**, *312*, 207–213. [[CrossRef](#)]
78. Yarham, J.W.; McFarland, R.; Taylor, R.W.; Elson, J.L. A proposed consensus panel of organisms for determining evolutionary conservation of mt-tRNA point mutations. *Mitochondrion* **2012**, *12*, 533–538. [[CrossRef](#)]
79. Jelesko, J.G.; Harper, R.; Furuya, M.; Gruissem, W. Rare germinal unequal crossing-over leading to recombinant gene formation and gene duplication in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 10302–10307. [[CrossRef](#)]
80. Xiao, H.; Jiang, N.; Schaffner, E.; Stockinger, E.J.; Van Der Knaap, E. A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* **2008**, *319*, 1527–1530. [[CrossRef](#)]
81. Rensing, S.A. Gene duplication as a driver of plant morphogenetic evolution. *Curr. Opin. Plant Biol.* **2014**, *17*, 43–48. [[CrossRef](#)] [[PubMed](#)]
82. Panchy, N.; Lehti-Shiu, M.; Shiu, S.-H. Evolution of gene duplication in plants. *Plant Physiol.* **2016**, *171*, 2294–2316. [[CrossRef](#)] [[PubMed](#)]